Biweekly Report 1

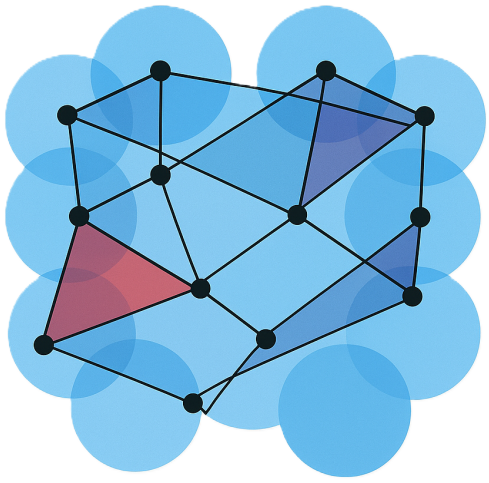
Using Persistent Homology to Compare Pathways and Discussion upon Creating New Species on the Local Machine

Raunak Narwal
Department of Mathematical Sciences
Indian Institute of Science Education and Research, Mohali, 130406, Punjab

August 7 to August 21, 2025

Persistent Homology

Complex structures like chemical reaction networks can be compared by persistent homology by capturing their topological features. A point cloud or network is given as input and we get a barcode or persistence diagram that shows which features persist across which scales as output. To apply persistent homology, we must convert chemical network into a mathematical object. We can compare two chemical networks by comparing their persistence diagrams using bottleneck distance and wasserstein distance.



Comparing Chemical Networks Using Persistent Homology

Chemical reaction networks can be represented as graphs, where metabolites (A substance made or used when the body breaks down food, drugs or chemicals, or its own tissue (for example, fat or muscle tissue)) form the nodes and enzymatic transformations form the edges. Traditional approaces to compare chemical networks focus on pathway overlaps or shared enzymes. These methods often miss out on global structural similarities that are critical to understanding functional equivalence between different biochemical systems. Persistent homology offers a way to adress this limitation by comparing the shape of chemical networks across multiple scales.

In this report, we are comparing two non trivial metabolic networks from KEGG databse using persistent homology. **Phenylalanine, tyrosine and tryptophan biosynthesis (KEGG map00400)** is an aromatic amino acid pathway present in bacteria, fungi and plants. **Terpenoid backbone biosynthesis (KEGG map00900)** is a pathway leading to precursors of steriods and carotenoids which is crucial for membrance stability and adaptation to stress.

These two pathways appear distinct at the biochemical level, their network topologies can be interrogated using persistent homology to reveal that whether they share organization such as recurring loops, feedback structures and modularity. By constructing Vietoris–Rips complexes from metabolite interaction graphs and comparing persistence diagrams using bottleneck and wasserstein distances, our aim is to determine whether these two pathways exihibit topological similatrity that could suggest analogous functional robustness.

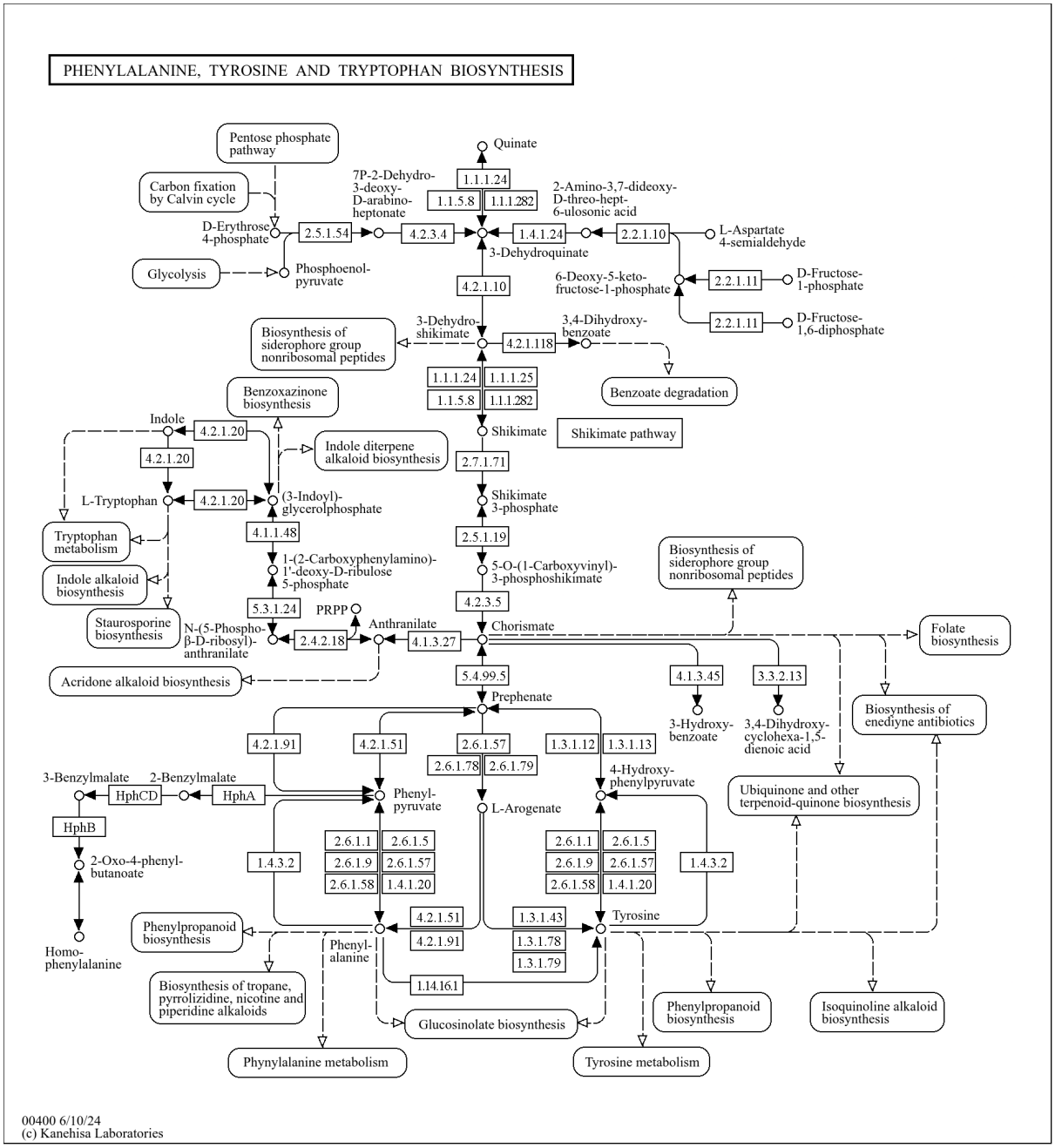


Figure 1: KEGG map00400

To compare the Phenylalanine, tyrosine and tryptophan biosynthesis pathway (KEGG map00400) with the Terpenoid backbone biosynthesis pathway (KEGG map00900), we used a computational pipeline based on persistent homology. Each pathaway is represented as a graph (network), the nodes of the graph are chemical compounds and the edges represent biochemical reactions connecting one compound to another. To apply topological methods, we require to have a notion of distance between compounds in the networks, we calculate the shortest path distances between all pairs of nodes, making a distance matrix. Using the distance matrix, we construct a Vietoris–Rips complex, which is a higher dimensional structure (intuitively) that encodes the connectivity of the network at different distance thresholds. As the threshold is increased, new connections and loops start to form in the complex. With the Rips complex, we compute persistent homology, which

tracks how topological features appear and disappear across scales. H_0 features represent connected components (clusters of compounds) and H_1 features represent loops or cycles in the reaction network. The results are visualized in a persistence diagram, where each point is a topological feature. To compare these pathways, we use distance metrics such as Wasserstein distance (measures the overall distributional difference between features in the two diagrams). If these distances are small, the pathways are similar; if large then they differ significantly. Bottleneck distance measures the largest difference between matched features in the two diagrams.

Demo run: Validating the Persistent Homology Pipeline

Before applying our method to chemical networks, we first validated the pipeline using a demo run. We were unable to obtain KGML files for the two above-mentioned pathways (KEGG Markup Language format). We thought to test our framework on synthethic graphs instead. This ensured that the implementation and computational steps were working as expected. In the demo run, we constructed two simple networks, `Cycle5`: a cycle graph with 5 nodes, which contains a clear topological loop and `Path5`: a path graph with 5 nodes, which is acyclic and therefore lacks loops.

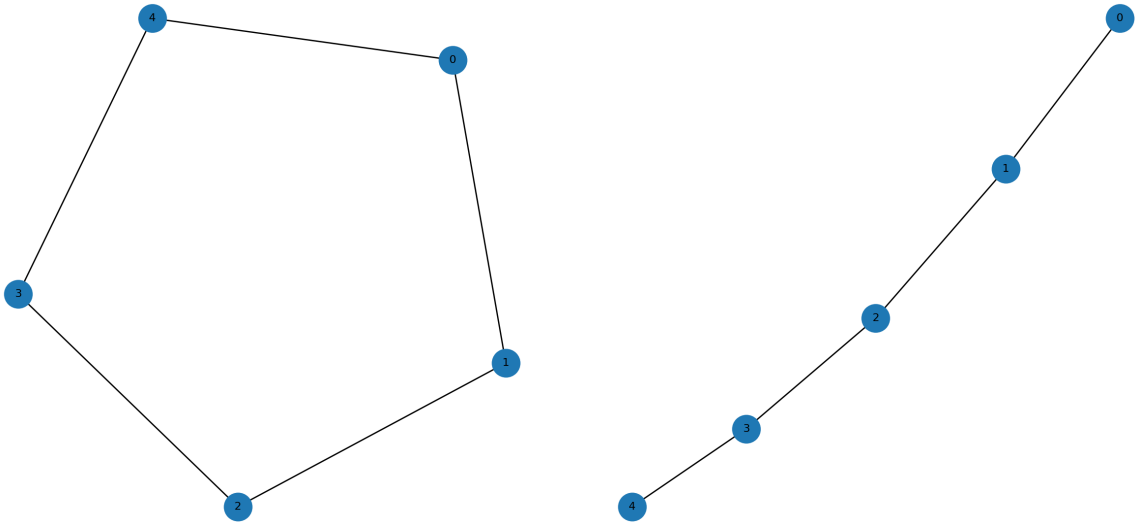


Figure 2: Left: Cycle5 (looped). Right: Path5 (acyclic).

Running the pipeline, we observed that the persistence diagram for Cycle5 clearly shows a non trivial H_1 (corresponding to the loop structure) and the persistence diagram for Path5 contains only H_0 features, reflecting connected components but no higher dimensional holes.

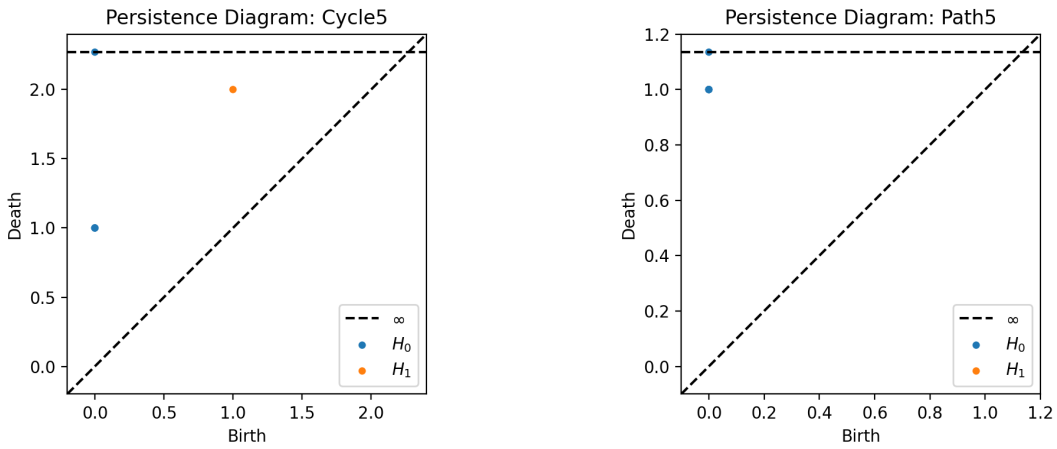


Figure 3: Persistence diagrams from demo run.

Plans for Next Period

Generated: August 19, 2025